DATABASES

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

# A New Face and New Challenges for Online Mendelian Inheritance in Man (OMIM®)

Joanna Amberger, Carol Bocchini, and Ada Hamosh*

*McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland*

For the HVP Bioinformatics Special Issue

Received 10 December 2010; accepted revised manuscript 19 January 2011.
Published online 5 April 2011 in Wiley Online Library (www.wiley.com/humanmutation). **DOI: 10.1002/humu.21466**

**ABSTRACT**: OMIM's task of cataloging the association between human phenotypes and their causative genes (the Morbid Map of the Genome) and classifying and naming newly recognized disorders is growing rapidly. Establishing the relationship between genotype and phenotype has become increasingly complex. New technologies such as genome-wide association studies (GWAS) and array comparative genomic hybridization (aCGH) define "risk alleles" that are inherently prone to substantial interpretation and modification. In addition, whole exome and genome sequencing are expected to result in many reports of new mendelian disorders and their causative genes. In preparation for the onslaught of new information, we have launched a new Website to allow a more comprehensive and structured view of the contents of OMIM and to improve interconnectivity with complementary clinical and basic science genetics resources. This article focuses on the content of OMIM, the process and intent of disease classification and nosology, and anticipated improvements in our new Website (http://www.omim.org).
Hum Mutat 32:564–567, 2011. © 2011 Wiley-Liss, Inc.

**KEY WORDS**: OMIM; nosology; disease classification; database

## Introduction

For over 40 years, OMIM has cataloged human mendelian disease and has focused on the relationship between genes and their molecular variants and associated phenotypes [McKusick, 2007]. The exponential growth of the database reflects the growth of knowledge in the field of medical genetics. As of 29 November 2010, OMIM had over 20,267 entries describing 13,606 genes and over 7,000 disorders. OMIM continues to add new gene descriptions to the database, with priority for creation given to genes with clinical relevance and functional significance. The phenotype entries in OMIM primarily describe single gene mendelian disorders, for example, sickle cell anemia (MIM# 603903), which is caused by mutation in the hemoglobin beta gene (*HBB*; MIM# 141900), and achondroplasia (MIM# 100800), which is caused by mutation in the fibroblast growth factor receptor 3 gene (*FGFR3*; MIM# 134934). Other phenotype entries describe complex traits for which variation in a single gene results in significant contribution to the phenotype, for example, macular degeneration (see MIM# 603075), skin/hair/eye pigment variation (see MIM# 227220), and inflammatory bowel disease (see MIM# 266600). A number of inherited disorders previously thought to be single gene disorders, such as Williams-Beuren syndrome (MIM# 194050), Smith-Magenis syndrome (MIM# 182290), and DiGeorge/velocardiofacial syndrome (e.g., MIM# 188400), have been shown to be contiguous gene deletion or duplication syndromes. Recent advances in chromosome microarray technology have allowed for rapid discovery of new members in this class of inherited genetic disease. This expanding category of copy number variation (CNV) disorders are now included in OMIM.

In OMIM, mutations known to cause a phenotype are cataloged in the allelic variants section of the causative gene entry. Only selected mutations are included, based on the following criteria: the first mutation to be discovered, a distinctive phenotype, an unusual type of mutation for a specific phenotype, an unusual pathogenetic mechanism, high population frequency, a distinctive inheritance (e.g., dominant with some mutations and recessive with others in the same gene), and historically important mutations. Most of the allelic variants represent disease-producing mutations; however, a few polymorphisms are included, many of which show a positive statistical correlation with particular common d isorders. As of 29 November 2010, OMIM had over 18,000 allelic variants distributed among 2,494 genes and associated with 4,218 different disorders or susceptibilities (Fig. 1). A tabular view of the allelic variant information is available from the "Table View" link under the Allelic Variant heading in gene entries and in the "Table View" link in the Table of Contents menu bar in gene entries. In addition, a summary of the gene phenotype relationships (the Morbid Anatomy of the Human Genome) is presented at the top of each entry (Fig. 2).

## Nosology

The naming and classification of mendelian phenotypes is essential to our understanding of biologic variation, and MIM has played a central role in this nosology. The process of disease classification involves defining recognizable patterns of features and highlighting those that allow one condition to be distinguished from another. Classifying disease is an evolving process that is affected by diagnostic modalities, medical intervention, molecular understanding, and community consensus. Sound nosologic practices will aid clinicians in diagnosis, prognosis, counseling, and management, and will aid researchers in elucidation of disease etiology.

*Correspondence to: Ada Hamosh, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21287.
E-mail: ahamosh@jhmi.edu

Contract grant sponsor: NHGRI; Contract grant number: 3UO1HG004438 (for OMIM curation and updating).

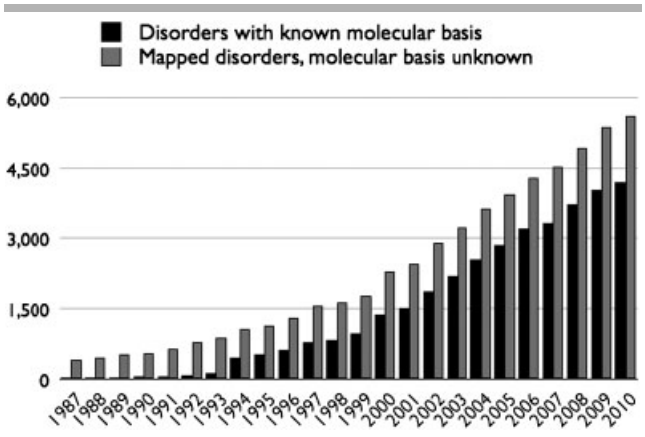


**Figure 1.** Growth of disease gene discovery 1987 to November 29, 2010.

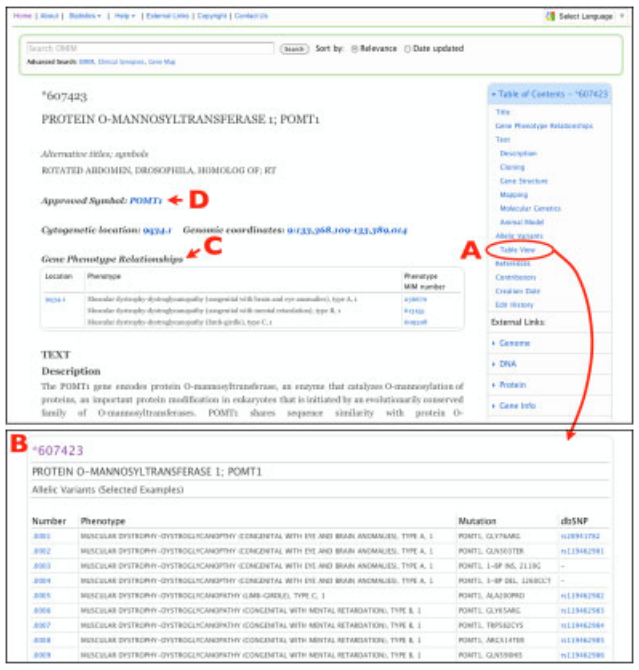


**Figure 2.** Screen capture of the entry for the *POMT1* gene in OMIM (http://omim.org/entry/607423). **A:** Link from entry to Allelic Variant table, **B, C:** Summary listing of Gene Phenotype relationships. **D:** Human Gene Nomenclature Committee (HGNC)-approved gene symbol with link to HGNC website. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]

The naming of a phenotype in OMIM is a complex, multistep process. Myriad initial questions must be answered; for example, what features actually define the phenotype? Does this constellation of features represent a new entity? Do different features of a disorder constitute clinical variability of a single disorder or define separate disorders? Have the same or similar features been described under a different name? Is the phenotype similar to others in OMIM? Can the phenotype be classified with any other disorders? Answering these questions must also take into account the views and possible disagreements in the clinical and genetic communities, with whom we are in frequent consultation, as well as published nosologies such as the IC3D classification of the corneal dystrophies [Weiss et al., 2008]. Finally, because the definition of a constellation of features as a genetic entity is an evolving process, the naming of disorders in OMIM may change over time. What does not change in OMIM is the unique identifier (MIM♯) assigned to each entity. MIM numbers are used widely in the literature and have proven to be useful, stable identifiers in the classification of human phenotypic variation.

Two major concepts of medical genetics, genetic (or locus) heterogeneity, and phenotypic diversity at a locus are basic to the naming of phenotypes. A genetically heterogeneous disorder is clinically (phenotypically) similar, but fundamentally (genotypically) distinct. This distinction is important not only for diagnostic purposes but also for therapeutic and pharmacologic intervention. In general, OMIM creates separate phenotype entries based on molecular etiology, that is, genetic heterogeneity. An example of the appropriateness of this approach is the naming of the long QT syndromes. Clinically, the disorder is characterized by a prolonged QT interval on electrocardiogram. However, pharmacologic intervention is based specifically on which ion channel or channels are mutated. The long QT syndromes caused by potassium ion channel mutations respond better to β-adrenergic blockade. The others may respond, but the likelihood of needing further intervention (such as an implantable defibrillator) is higher. OMIM now has 13 long QT syndrome entries, each with a unique molecular basis. The lowest number in the series, LQT1 (MIM♯ 192500), contains summary information on the topic.

In phenotypic diversity at a locus, different mutations in one gene cause several different disorders. For example, mutations in lamin A (LMNA; MIM# 150330) can cause 11 different conditions, including Emery-Dreifuss muscular dystrophy (MIM# 181350), mandibuloacral dysplasia with type A lipodystrophy (MIM# 248370), and Hutchinson-Guilford progeria (MIM# 176670). Similarly, mutations in *FGFR3* (MIM# 134934) can cause eight disorders, including achondroplasia (MIM# 100800), lacrimoauriculodentodigital syndrome (LADD; MIM# 149730), and lethal thanatophoric dysplasia (TD1; MIM# 187600). Although some of the disorders caused by variation in the same gene are clearly distinct, others overlap in features or severity and require further nosologic clarification.

Sometimes what appears to be phenotypic diversity at a locus turns out, in fact, to be a single clinical entity. What was recently described as a "new Ehlers-Danlos syndrome with craniofacial characteristics, multiple congenital contractures, progressive joint and skin laxity, and multisystem fragility-related manifestations" [Kosho et al., 2010] was determined to have similar features to an existing entity in OMIM, adducted thumb–clubfoot syndrome (ATCS; MIM# 601776), which is due to mutations in *CHST14* (MIM# 608429). Some of the patients reported by Kosho et al. [2010] had been reported earlier [Kosho et al., 2005] as having Ehlers-Danlos syndrome VIb (EDS VIB). EDS VIB had also been used to describe a different disorder now known as brittle cornea syndrome (MIM# 229200). Malfait et al. [2010] reported additional patients with *CHST14* mutations and confirmed the overlap of ATCS and some of the EDS VIb patients. The name for the phenotype characterized by *CHST14* mutations is "Ehlers-Danlos syndrome, musculocontractural type." Interestingly, this example highlights the value of reports of long-term follow-up in the understanding of disease manifestations and progression and the classification of disease. In the original description of adducted thumb–clubfoot syndrome by Dundar et al. [1997], the patients were very young, and joint contractures were the most prominent features. The patients reported by Kosho et al. [2010] were older, and skin and joint laxity were the most prominent features.

The task of defining the constellation of features of a disease entity may be confounded by several factors. Many genetic

disorders are rare and case reports may be published years apart. In addition, the subspecialization of medicine often creates a "blind men and the elephant" scenario because each specialty focuses on the manifestations of the disorder in a specific organ system. This is particularly true in mendelian disorders, many of which exhibit pleiotropy or manifestations in many seemingly unrelated organ systems due to the effects of the same dysfunctional protein. Marfan syndrome (MIM# 154700), for example, was originally characterized as a skeletal disorder called dolichostenomelia [Marfan, 1896]. With time, it became evident that cardiac and ocular findings were cardinal features of the disorder. Complex disorders present with similar confounding problems. For example, nephropathy, retinopathy, and peripheral vascular disease are all possible complications of diabetes and have been studied by different clinical subspecialities, with their results appearing in different subspeciality journals. OMIM recognized the common underlying clinical principle when classifying them in a phenotypic series designated "microvascular complications of diabetes" (see MVCD1; MIM# 603933). This classification provided a logical framework to map similar genomic findings from various genome-wide association studies of different (renal, retinal, peripheral disease) patient cohorts.

Unlike model organisms, in which the full range of phenotypic expression is open to ascertainment by investigators, the characterization of human disorders presents many practical problems. Our ability to accurately phenotype a human is based on the availability of patients and samples, which are limited due to patient safety and ethical and cultural considerations, and the availability of technology. For example, the advent of amino acid chromatography and of gas chromatography allowed the recognition of aminoacidopathies and organic acidemias, respectively. The development of isoelectric focusing of transferrin and mass spectrometric methods were required for the classification of the now over 25 congenital disorders of glycosylation (see MIM# 212065). Improved patient imaging technologies, which allow for earlier diagnosis, have facilitated the recognition of important clinical complications and refined phenotyping by highlighting unifying features among previously unrelated disorders. An example is the widening group of syndromes characterized by arterial tortuosity (e.g., the Loeys-Dietz syndromes [MIM#s 610168, 609192, 608967, and 610380], arterial tortuousity syndrome [MIM# 208050], and some forms of cutis laxa [MIM# 219100]. The similarity among these disorders could not be recognized before the advent of full body angiography by either computed tomography or magnetic resonance. This new classification is of critical clinical importance when considering surgical interventions.

It is sometimes necessary to reclassify phenotypes. Because the practice of medicine is to intervene in and ameliorate human disease, the natural history of a disorder may change. The change may result in the diminution or disappearance of a major feature of a disorder, or patients may survive much longer, thereby revealing additional clinical complications. For example, congenital hypothyroidism had been classified as "cretinism" based on the rapidly progressive neurologic, cognitive, and motor damage seen in untreated infants and children. Now new cases of congenital hypothryoidism are promptly diagnosed and treated, thereby eliminating that feature from the disorder. Currently, there are 15 different congenital hypothyroidism entities in OMIM.

Human phenotypes may also be reclassified based on a unifying principle derived from the molecular pathway or biology underlying them. An example of this involves the categorization of a subset of the broad category of congenital muscular dystrophies that were found to share a defect in any one of six genes in the glycosylation pathway of alpha-dystroglycan (*DAG1*; MIM# 128239). Many of these disorders had previously been classified separately or lumped together regardless of molecular basis, for example, Walker-Warburg syndrome (WWS), muscle–eye–brain disease (MEB), Fukuyama congenital muscular dystrophy (FCMD), and limb–girdle muscular dystrophy type 2K (LGMD2K). The discussion of these disorders in the literature as dystroglycanopathies [Godfrey et al., 2007] led to the creation in OMIM of a muscular dystrophy-dystroglycanopathy (MDDG) series (Table 1). The most severe form of the disorder, which has brain and eye anomalies, severe disability and/or death at an early age, was previously designated WWS, MEB, or FCMD. In the updated classification, it is designated type A (MDDGA). An intermediate form, with or without mental retardation, previously called simply congenital muscular dystrophy, is now designated type B (MDDGB). The least severe form, characterized by later onset and no mental retardation and previously designated limb–girdle muscular dystrophy or congenital muscular dystrophy, is now designated type C (MDDGC). Each type is further classified and numbered based on the molecular defect, for example, MDDGA1 (MIM# 236670), MDDGB1 (MIM# 613155), and MDDGC1 (MIM# 609308) are all caused by mutation in the *POMT1* gene (MIM# 607423); MDDGA2 (MIM# 613150), MDDGB2 (MIM# 613156), and MDDGC2 (MIM# 613158) are all caused by mutation in the *POMT2* gene (MIM# 607439), etc. Previous designations for all of these disorders are included as alternative titles in the appropriate entries. The updated classification provides discrete entries in which to discuss the intricacies of the genotype–phenotype relationship, and a tabular view of the phenotypic series is available from the phenotype entries (Fig. 3).

OMIM is part of a worldwide effort to improve disease classification. These efforts include establishing a standard nomenclature for features of a disorder (traits) (e.g., the Human Phenotype Ontology) and coding broad disease classifications for billing, public health, and outcomes (e.g., the International Classification of Diseases and HL7 initiatives). OMIM has a unique long-term perspective on naming diseases with regard to their molecular origins and is responsible for classifying and naming human mendelian disease.

**Table 1. Clinical Classification of Muscular Dystrophy-Dystroglycanopathies (MDDG)**

| MDDG type | Clinical findings |
|---|---|
| Type A | |
| *includes Walker-Warburg syndrome, Muscle-Eye-Brain disease, and Fukuyama MD* | Structural brain anomalies<br>Structural brain anomalies<br>Ocular anomalies<br>Congenital hypotonia, severe<br>Sitting and walking not achieved<br>Severe global developmental delay and mental retardation<br>Decreased lifespan |
| Type B | |
| *includes congenital muscular dystrophy* | No No major structural brain anomalies<br>No major eye anomalies<br>Congenital hypotonia<br>Sitting and walking may be achieved<br>Mental retardation (in most cases) |
| Type C | |
| *includes limb-girdle muscular dystrophies 2I, 2K, 2M, 2N* | Later onset<br>Proximal muscle weakness<br>Difficulty walking<br>Normal cognition (in most cases) |

*Note*: The genes involved in muscular dystrophy-dystroglycanopathy are (1) POMT1, (2) POMT2, (3) PMGNT1, (4) FKTN, (5) FKRP, and (6) LARGE.

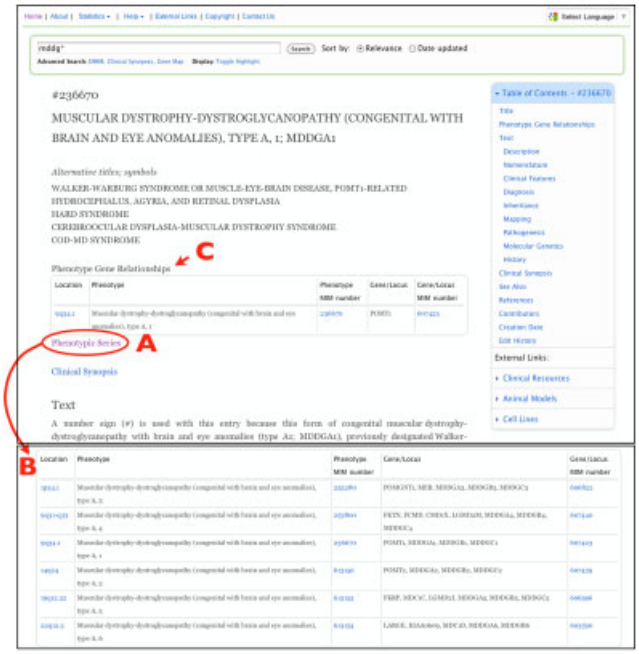


**Figure 3.** Screen capture of the entry for muscular dystrophy-dystroglycanopathy type A1 (MDDGA1) (http://omim.org/entry/236670). **A:** Phenotypic series link to **B**, a table of the phenotypic series of muscular dystrophy-dystroglycanopathy type A. **C:** A tabular view of the phenotype gene relationship of MDDGA1. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]

## OMIM's New Website

In late 2010, OMIM launched a beta version of a new Website at http://www.omim.org. This is the first major overhaul of its Web presence since 1999. The intent of the Website is to create a clinical portal to the genome with OMIM as the central resource from which users can easily and quickly navigate to related information. We are working on an advanced search function that will include the option of layering a thesaurus into a search query. This will facilitate retrieval of similar concepts such as mental retardation, developmental disability, and intellectual disability. To further aid in searching clinical features, users will be able to restrict their searches to major anatomical headings within the Clinical Synopses. We will add links from the features in each OMIM clinical synopsis to more structured ontologies, such as the Human Phenotype Ontology [Robinson et al., 2008], the Elements of Morphology Terms [Allanson et al., 2009], ICD10, and the Mammalian Phenotype Ontology [Smith et al., 2005].

Links in the new Website will be organized by category (genomic, clinical, variation, etc.) and made readily accessible to the user. In addition to links to GeneTests, GeneReviews, and the Genetic Alliance, we will be creating links to clinical images resources such as POSSUM and DermAtlas, rare disease resources such as NORD and OrphaNet, and clinical management tools such as clinicaltrials.gov and newborn screening resources. We are collaborating with the genome browser providers to provide views of their information that have a clinical genetics perspective.

## Collaboration

OMIM is interested in collaborating with anyone who would like to mine the resource. We encourage people to contact us directly so that we can provide guidance to mining the information most effectively. A "Contact Us" link is available at the top of all OMIM pages to facilitate input from the community. OMIM is freely accessible for individual use and will remain so, but reasonable fair-use licenses are requested for full data downloads.

## Summary

OMIM focuses on rare diseases that have a significant genetic basis. As a resource based exclusively on the biomedical literature and with a team of highly expert curators, we provide authoritative, comprehensive, and timely descriptions with links to genomic, model organism, and other research resources, as well as to variation, coding (e.g., ICD-10), clinical trials, and other clinical resources. We serve two separate and overlapping communities: molecular biologists and healthcare providers caring for patients with mendelian disorders, as well as students of both disciplines. Today, as the use of whole genome and exome sequencing rapidly advances our knowledge of mendelian disorders, OMIM's role of classifying these disorders and the biological variation underlying them has never been more important.


## Acknowledgments

OMIM's new Website is funded by The Johns Hopkins University and a grant for the Maryland Department of Health and Mental Hygiene.


## References


Allanson JE, Biesecker LG, Carey JC, Hennekam RC. 2009. Elements of morphology: introduction. Am J Med Genet 149A:2–5.

Dundar M, Demiryilmaz F, Demiryilmaz I, Kumandas S, Erkilic K, Kendirci M, Tuncel M, Ozyazgan I, Tolmie JL. 1997. An autosomal recessive adducted thumb–club foot syndrome observed in Turkish cousins. Clin Genet 51:61–64.

Godfrey C, Clement E, Mein R, Brockington M, Smith J, Talim B, Straub V, Robb S, Quinlivan R, Feng L, Jimenez-Mallebrera C, Mercuri E, Manzur AY, Kinali M, Torelli S, Brown SC, Sewry CA, Bushby K, Topaloglu H, North K, Abbs S, Muntoni F. 2007. Refining genotype–phenotype correlations in muscular dystrophies with defective glycosylation of dystroglycan. Brain 130:2725–2735.

Kosho T, Miyake N, Hatamochi A, Takahashi J, Kato H, Miyahara T, Igawa Y, Yasui H, Ishida T, Ono K, Kosuda T, Inoue A, Kohyama M, Hattori T, Ohashi H, Nishimura G, Kawamura R, Wakui K, Fukushima Y, Matsumoto N. 2010. A new Ehlers-Danlos syndrome with craniofacial characteristics, multiple congenital contractures, progressive joint and skin laxity, and multisystem fragility-related manifestations. Am J Med Genet Part A 152A:1333–1346.

Kosho T, Takahashi J, Ohashi H, Nishimura G, Kato H, Fukushima Y. 2005. Ehlers-Danlos syndrome type VIB with characteristic facies, decreased curvatures of the spinal column, and joint contractures in two unrelated girls. Am J Med Genet Part A 138A:282–287.

Malfait F, Syx D, Vlummens P, Symoens S, Nampoothiri S, Hermanns-Le T, Van Laer L, De Paepe A. 2010. Musculocontractural Ehlers-Danlos syndrome (former EDS type VIB) and adducted thumb clubfoot syndrome (ATCS) represent a single clinical entity caused by mutations in the dermatan-4-sulfotransferase 1 encoding CHST14 gene. Hum Mutat 31:1233–1239.

Marfan AB. 1896. Un cas de deformation congenitale des quatre membres, plus prononcee aux extremites, caracterisee par l'allongement des os avec un certain degre d'amincissement. Bull Mem Soc Med Hop Paris 13:220–226.

McKusick VA. 2007. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet 80:588–604.

Robinson PHN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 83:610–615.

Smith CL, Goldsmith CA, Eppig JT. 2005. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biol 6:R7.

Weiss JS, Moller HU, Lisch W, Kinoshita S, Aldave AJ, Belin MW, Kivela T, Busin M, Munier FL, Seitz B, Sutphin J, Bredrup C, Mannis MJ, Rapuano CJ, Van Rij G, Kim EK, Klintworth GK. 2008. The IC3D classification of the corneal dystrophies. Cornea 27(Suppl 2):S1–S42.